

July 5, 1947.

Dear Dr. Rhoades-

How is everything?

You may recall that I mentioned in my seminar on coli linkage that for most of my map, I could give only relative, not absolute distances. For some time now, I have been trying to develop a theory to estimate the ~~what~~ absolute distance involved by means of the frequency of the triple-exchange types. Briefly the situation is this:  $\frac{ABcD}{ABCD}$ , and call the regions between and b, a, etc.

The only types recovered (as prototrophs) are:

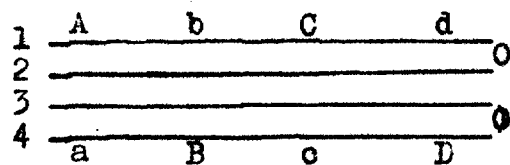
$\frac{ABcD}{ABcD}$  27.3%     $\frac{AbcD}{AbcD}$  46.3%     $\frac{AbcD}{AbcD}$  24.3% , which are exchanges in regions a, b and c respectively, and  $\frac{ABCD}{ABCD}$  2.1% which is a triple-exchange type.

The theory relating the frequency of  $\frac{ABCD}{ABCD}$  to the absolute distances, on a two-strand theory, is very simple, and barring interference, a map of 72 morgans between A and D is obtained. I am still trying to develop a rigorous theory for a similar estimate on the four-strand theory, with only beginning success, because of the complexity of the problem when dealing with 3, 4, and 5 random crossovers per tetrad. My first approximate solution is about 80 morgans, which is surprisingly close to the two-strand figure. Do you know of any theoretical treatment of multiple crossing over that would facilitate my labors?

In the course of this armchair experimentation (which is a good deal more laborious than lab.-desk work!), I developed an "Operator" notation and method of dealing with multiple-crossing over, concerning which I should like to hear your opinion relative to a) its adequacy b) its usefulness in general and c) its novelty. The method is a good deal easier to use than to explain; it concerns the problem of cal-

culating the nature of all the different tetrads which result from combinations of various crossovers:

Write:



Each crossover is written with the region involved, and the strands exchanged as subscripts, e.g.  $a_{13}$ ,  $b_{24}$ , etc. The crossovers in a given tetrad are written in sequence, and may be unlimited in number.

Each such symbol is regarded as an operator, which affects only that strand bearing the same subscript as it does (i.e. of the same rank). The effect of a c.-o. operation  $xy$ , going from right to left is a) to substitute certain alleles by their alternative, depending on the region of the crossover, i.e.

$a_{xy}$  substitutes a/A and A/a, written (a), operating on strands of rank x and y.

$b_{xy}$  substitutes a/A; b/B and vice versa, .... (ab)

$c_{xy}$  substitutes a/A; b/B and c/C, and vice versa, .....(abc).

A second effect of each operation, e.g.  $a_{xy}$  on  $s_x$  is to change its rank from x to y with respect to any further operators.

This, is, of course, merely a restatement of the interaction of crossovers, but the symbolic formulation is useful, since one can write down certain combinations at once:

$$\begin{aligned} a.b &= (a).(ab) = (b) & abc &= a.bc = (a)(ac) = (c). \\ a.c &= (a).(abc) = (bc) \\ b.c &= (b).(abc) = (ac) \end{aligned}$$

E.G., if s is ABCD,

$ab.s = \cancel{ab} s = AbCD$	$abc.s = ABcD$
$ac.s = AbcD$	$a.s = aBCD$
$bc.s = AbcD$	$b.s = abCD$
$aa.s = bb.s = cc.s = ABCD$	$c.s = abcd.$

Obviously the repetition of an operator restores the original configuration.

To break down a sequence of crossovers into its component operations on each strand, one starts from right to left, and writes the appropriate operation, using the subscript to indicate the new rank, until the c.o.s are exhausted. E.G.,  $a_{13}.a_{14}.b_{23}.c_{24}$  is:

On  $s_1$  the right handmost operator is  $a_{14}$ . One writes  $a_4.s_1$  or  $abCd$

On  $s_2$ , we see  $c_{24}$  on the righthand end, and write  $c_4.s_2$ . Looking for additional operators of rank 4, now, we see  $a_{14}$ , and write  $a_1c_4.s_2$ . Finally,  $a_{13}$  is seen to be an operator of rank 1, so we write  $a_3a_1c_4.s_2$ . This is, of course,  $ABcD$ . For  $s_3$ , we see  $b_{23}$ , and then

no additional rank 2 operators.  $b_2-s_3$  is  $AbcD$ . Finally, For  $s_4$  we write (from right to left):  $c_2$  then,  $b_3$ , then  $a_1$ ., i.e.  $a_1b_3c_2-s_4$  or  $ABCD$ . The tetrad is then:

abCd  
aBcD  
AbcD  
ABCD

For most purposes, the final subscripts can be neglected, except that 1,2 and 3,4 are the centromere markers, at either end you choose. The extension of this system to polysomic segregations is, I think, evident. After 5 minutes practice, the tetrads can be written down by inspection, instead of having to construct involved diagrams of the chromatids, particularly with tri- and quadrivalent associations. I am busy now trying to work out derived rules to see if the manipulation can be dispensed with altogether. Whether or not this system can be adapted to loop- and ring- formations, in aberration heterozygotes is another possibility I have not yet looked into.

Your opinion on all this would be greatly appreciated. It is hard to believe that no one else has done this sort of thing before, but I have not yet encountered it.

Best regards,

Yours sincerely,

Joshua Lederberg